ICCIT 2025
28th INTERNATIONAL CONFERENCE ON COMPUTER AND INFORMATION TECHNOLOGY
Long Beach Hotel
Cox's Bazar, Bangladesh
DECEMBER 19 - 21, 2025
http://iccit.org.bd/2025/
Organized By
IEEE Bangladesh Section

Paper ID: 1904

A Multifaceted Approach to Bangla Speech Emotion Recognition Using Hybrid Feature Fusion and Dimensionality Reduction

Abu Omayed, MD. Asif Iqbal, Mohammed Moniruzzaman Bhuiyan

University of Chittagong, Chittagong University of Engineering and Technology

Presented By: Abu Omayed

ICCIT COX'S BAZAR
28th International
Conference on Computer
and Information Technology
2025

Presentation Outline

1. Introduction
2. Motivation and Objectives
3. Methodology
4. Dataset
5. Performance Evaluation
6. Comparison
7. Conclusion
8. Reference

INTRODUCTION

- **Speech Emotion Recognition (SER) is critical for Human-Computer Interaction (HCI), but robust SER systems for the Bangla language remain underexplored.**

- **Speech is one of the most important means to express human emotions [1]. Emotions can reflect human psychological conditions (e.g. mood, anxiety, trauma) [2].**

- **Machine learning offers a promising approach to automatically Emotion recognition from Speech.**

OBJECTIVE

- **Analyse the correlation between feature selection and dimensionality reduction and their impact on the performance of the SER models.**
- **Introduce a unique multifaceted dataset by using a reduction approach to identify SER's most effective feature combinations.**
- **Evaluate multiple machine learning classifiers along with hyperparameter tuning to determine the best performing SER model.**

METHODOLOGY

Feature

- Amplitude Envelop
- RMS
- ZCR
- Mel Spectrogram
- MFCC
- Band Energy Ratio
- Spectral Centroid
- Spectral Bandwith

Statistics

- Mean
- Median
- Minimum
- Maximum
- Std.

PCA | **ICA** | **LDA**

Multifaceted Dataset

ML Model

- Decision Tree
- KNN
- Random Forest
- XGBoost

Evaluation

- Accuracy

Fig 1. Methodology of our Research

DATASET

Attribute	SUBESCO	KBES
Speech Source	Acted in Studio	TV/Film Media
Used Language	Standard Bangla	Standard Bangla
Sampling Rate	48 kHz	48 kHz
Number of Speakers	10 males, 10 females	15 males, 20 females
Age Groups	20–30, 30–40	Not Specified
Number of Emotions	7	9
Emotion States	Neutral, Angry, Disgust, Fear, Happy, Sad, and Surprise	Neutral, Angry, Disgust, Happy, Sad, Very Angry, Very Disgust, Very Happy, Very Sad
Number of Sentences	10	Not Applicable
Total Audio Clips	7000	900

TABLE I: Details of Used Datasets

MULTIFACETED DATASET

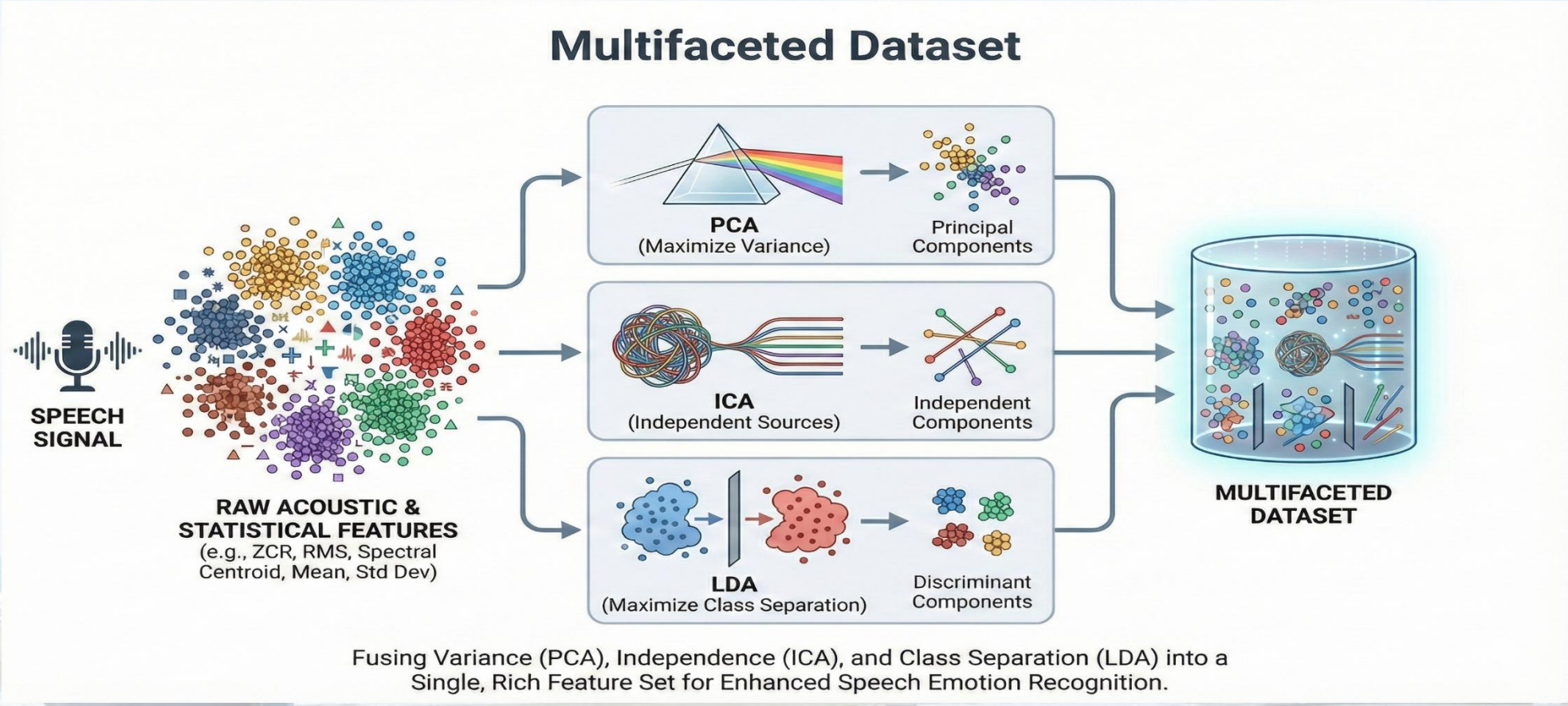


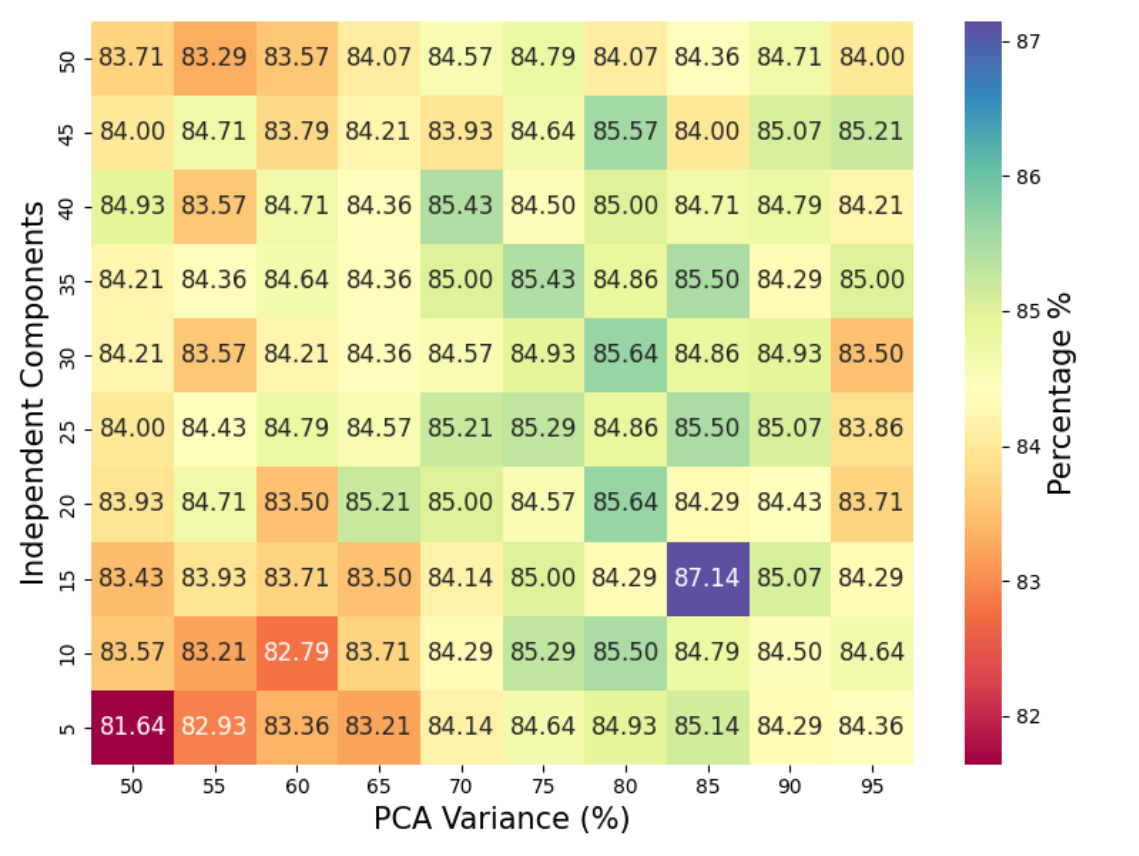
Fig. 2: The Multifaceted Feature Fusion Pipeline

HYPERPARAMETER

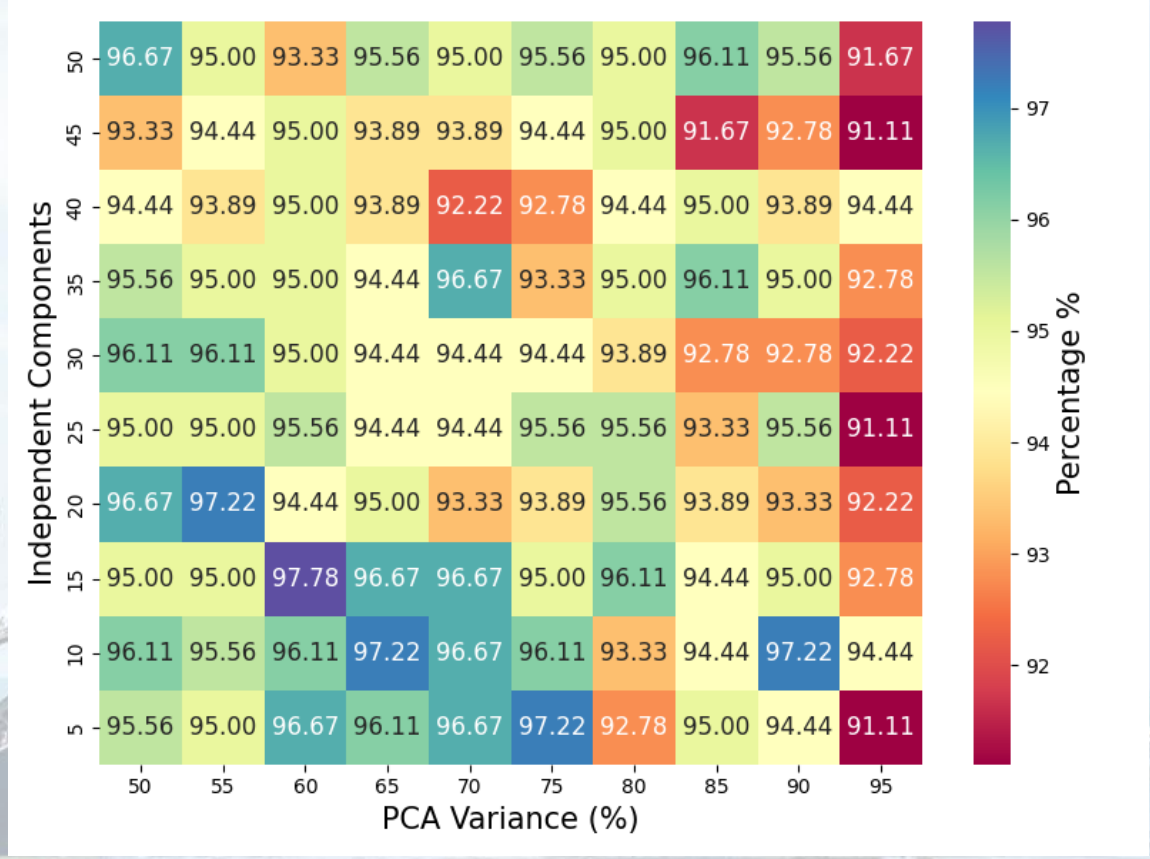
Proposed Model	Hyperparameter	Value
KNN	N_neighbors Metrics	7 Mankowski
DT	Max_depth	6
RF	N_etimatores	100
XGBoost	N_estimator Max_depth Learning_rate Objective	200 6 0.1 Softmax

TABLE II: Finalized Hyperparameters Used in Classification

MUTIFACETED DATSET PERFORMANCE



SUBESCO



KBES

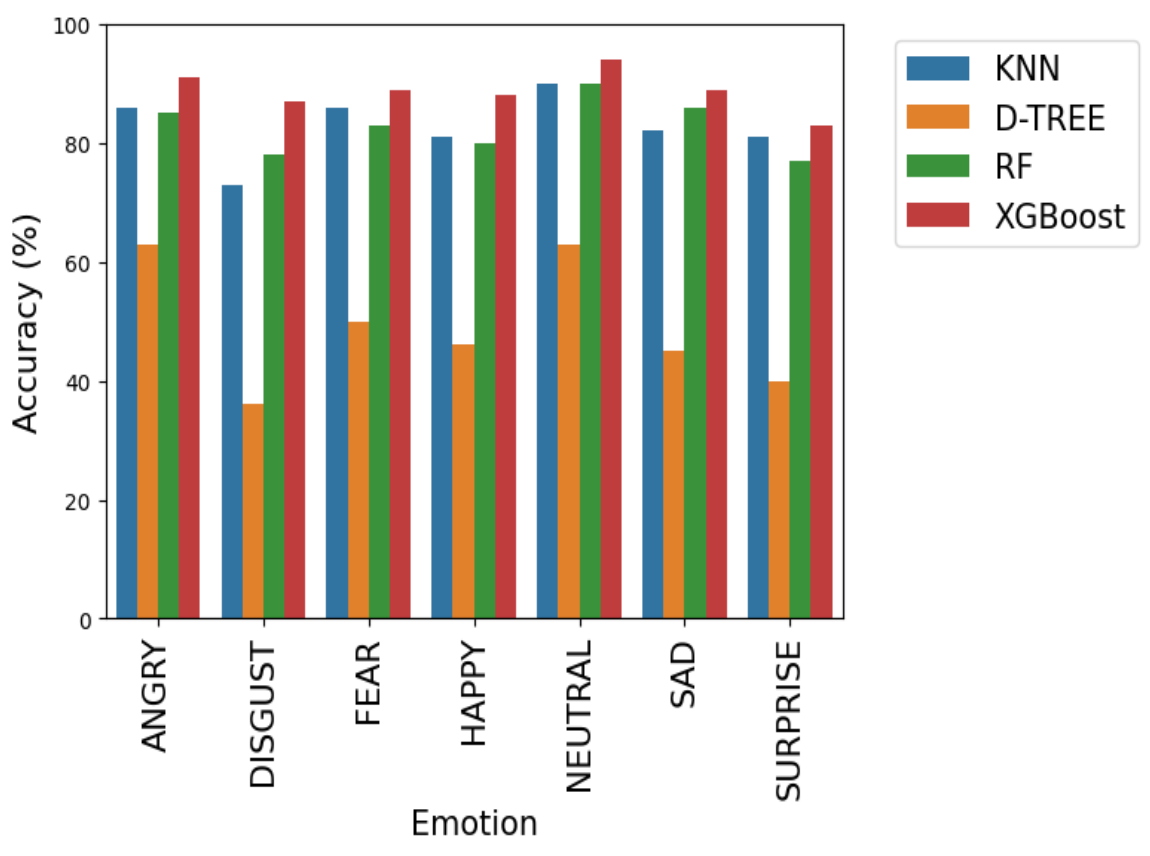
Fig. 2: Accuracy heatmap of RF for SUBESCO and KBES, PCA variation vs ICA components with LDA constant

MUTIFACETED DATSET PERFORMANCE

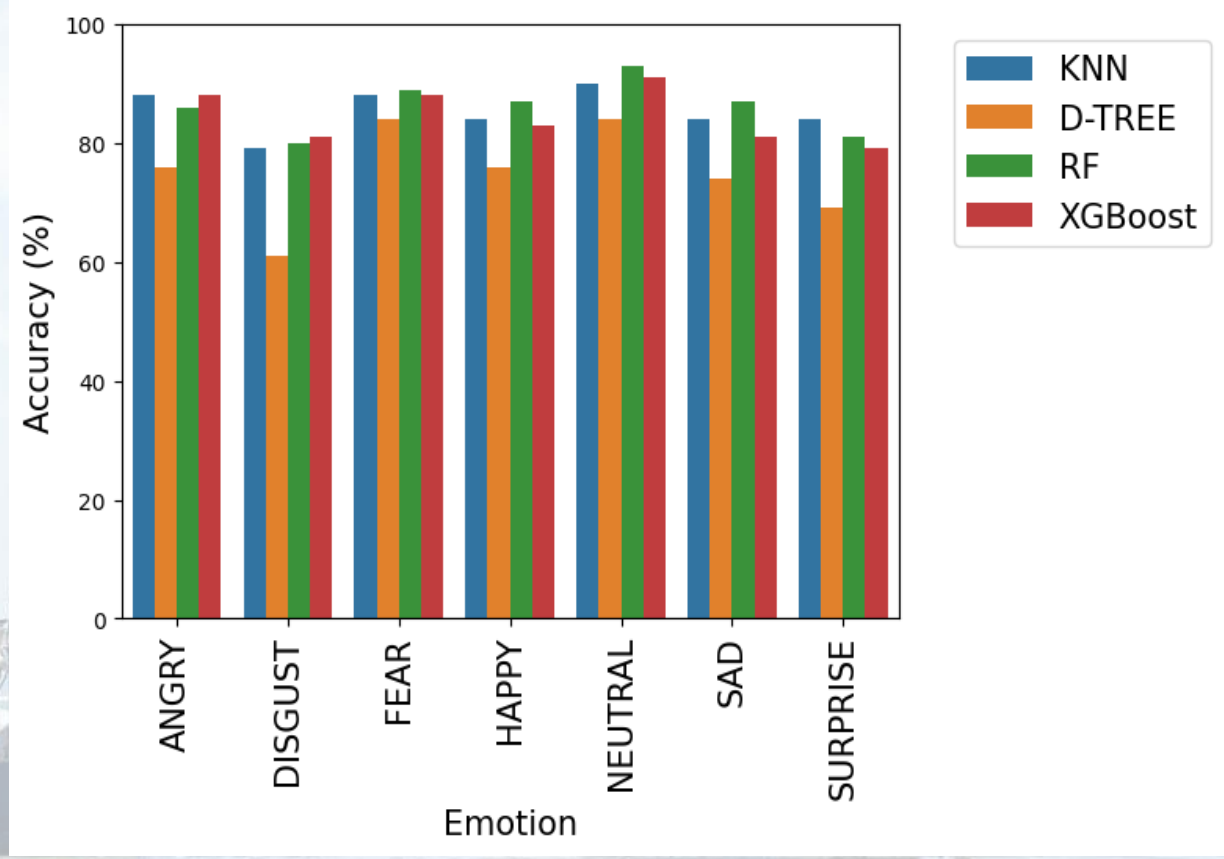
Dataset	Audio Features	Summary Statistics	Reduction Method	Classifier	# Features	Accuracy (%)
SUBESCO	Amplitude Envelope, RMS, ZCR, Mel-Spectrogram, MFCC, Band Energy Ratio, Spectral Centroid, Spectral Bandwidth	Mean, Median, Min, Max, Std.	None	KNN	430	82.24
			None	DT	430	48.13
			None	RF	430	82.93
			None	XGBoost	430	89.06
			PCA+ICA+LDA	KNN	137	85.21
			PCA+ICA+LDA	DT	78	74.93
			PCA+ICA+LDA	RF	86	**87.14**
			PCA+ICA+LDA	XGBoost	43	84.50
KBES	Same as SUBESCO	Mean, Median, Min, Max, Std.	None	KNN	430	50.89
			None	DT	430	34.78
			None	RF	430	59.64
			None	XGBoost	430	63.84
			PCA+ICA+LDA	KNN	23	73.89
			PCA+ICA+LDA	DT	86	80.00
			PCA+ICA+LDA	RF	40	**97.78**
			PCA+ICA+LDA	XGBoost	41	93.89

TABLE III: Comparison of Classifier Accuracies Using Feature Fusion and Dimensionality Reduction Techniques

FERFORMANCE EVALUATION



Without Data Reduction

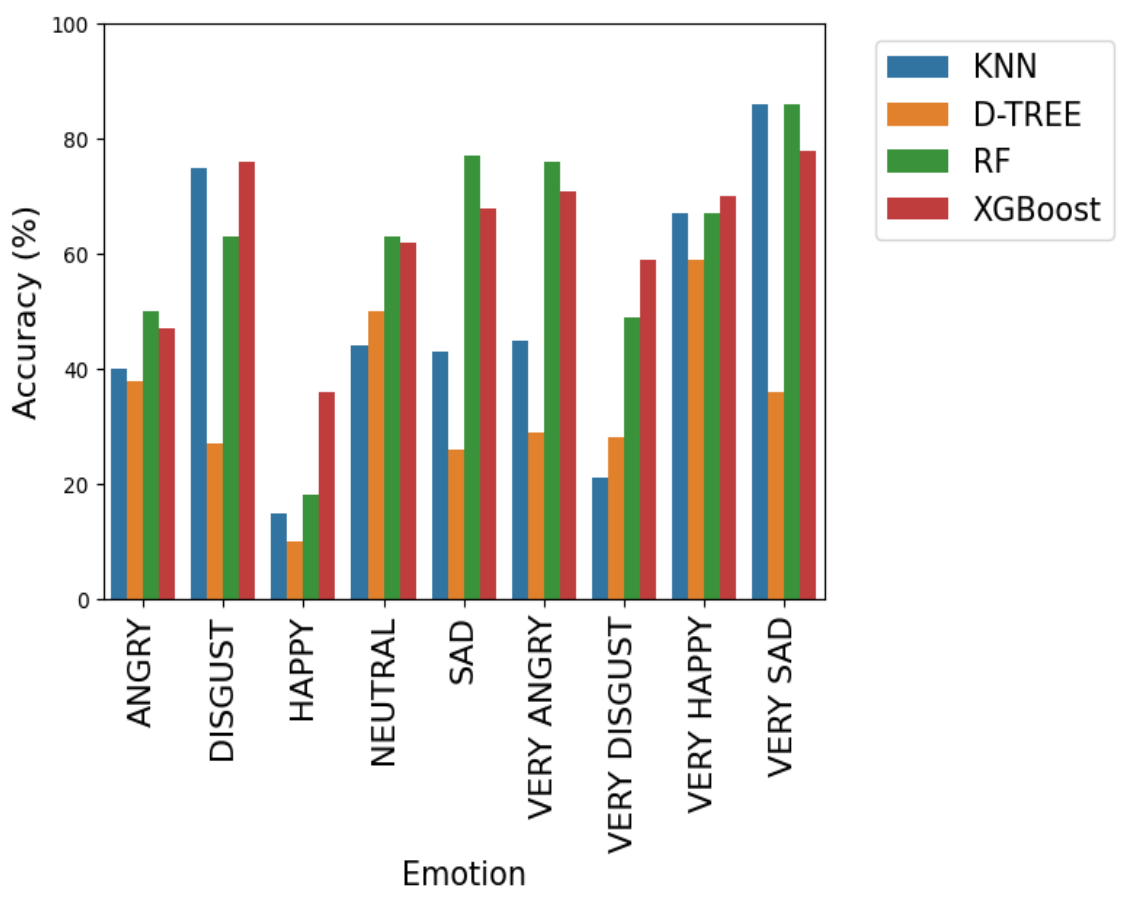


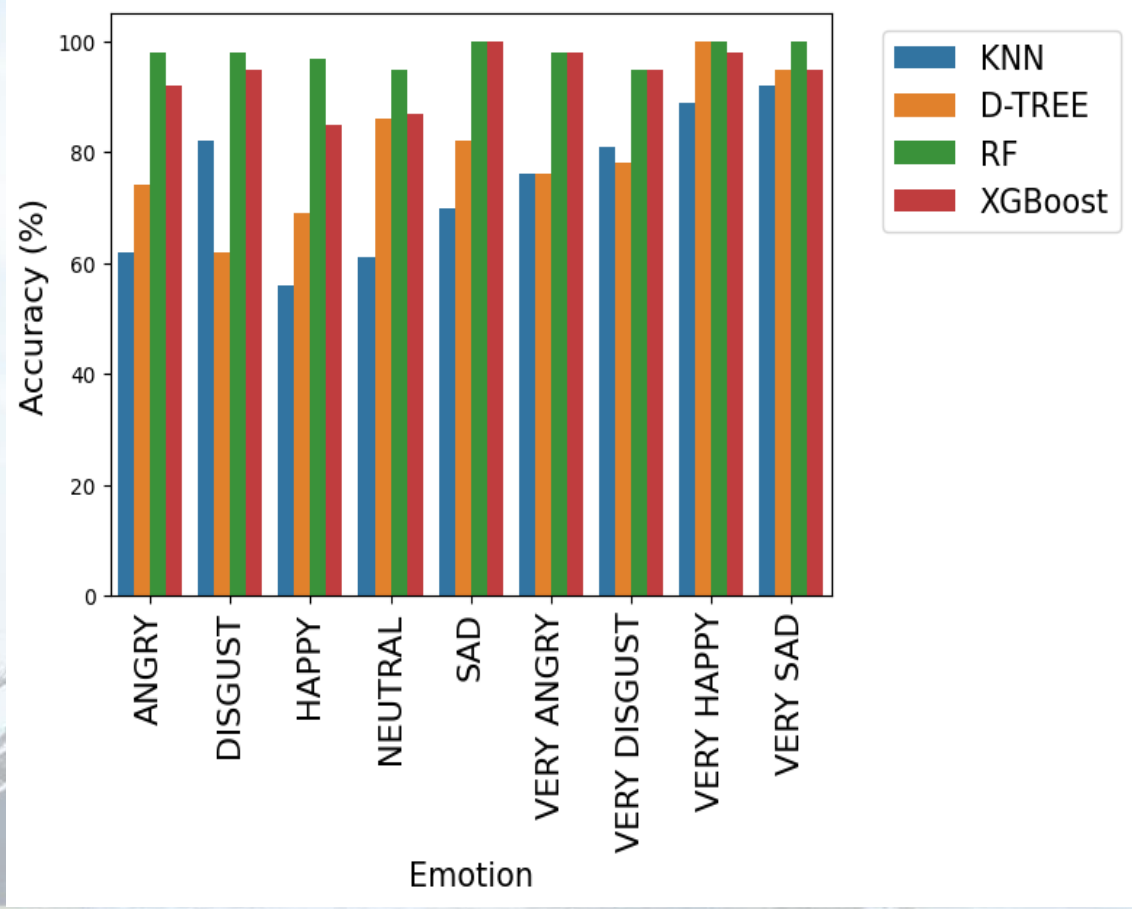
With Data Reduction (Multifaceted)

Fig. 3: Emotion recognition Performance by classifiers, (a) Without (b)With, Data Reduction techniques for SUBESCO

ICCIT COX'S BAZAR
28th International Conference on Computer and Information Technology
2025

FERFORMANCE EVALUATION





Without Data Reduction

With Data Reduction (Multifaceted)

Fig. 3: Emotion recognition Performance by classifiers, (a) Without (b)With, Data Reduction techniques for KBES

ICCIT COX'S BAZAR
28th International Conference on Computer and Information Technology
2025

COMPARISON

Author	Year	Dataset	Features	Data Reduction	Best Model	Accuracy
Bachchu Paul et al. [3]	2024	SUBESCO	MFCC, LPC, ZCR. Pitch, Energy	-	SVM	78.33%
Sadia Sultana et al. [4]	2024	SUBESCO	MFCC	-	Deep CNN	86.90%
Md. Masum Billah et al. [5]	2024	KBES	MFCC, STFT, CSTFT	-	Cascaded Deep Learning	71.67%
Proposed Method	-	SUBESCO & KBES	Amplitude Envelop, RMS, ZCR, Mel-Spectrogram, MFCC, BER	PCA, ICA and LDA	RF	87.14% and 97.78%

TABLE IV: Comparison with Related Works in SER

CONCLUSION

- **This study investigates the efficiency of combining feature fusion, statistical operations, and dimensionality reduction techniques to enhance Bangla SER.**

- **Our Multifaceted dataset have demonstrated their crucial role in enhancing the accuracy of machine learning classifiers. By using Random Forest classifier, we achieve 87.14% accuracy for SUBESCO dataset and 97.78% accuracy for KBES dataset.**

- **we plan to enhance the Bangla SER study by developing a new Bangla speech dataset to benefit the community.**

REFERENCE

1. P. Khanna and M. Sasikumar, “Recognizing emotions from human speech,” in Thinkquest˜ 2010: Proceedings of the First International Conference on Contours of Computing Technology, pp. 219–223,Springer, 2011.
2. K. A. Winter and N. A. Kuiper, “Individual differences in the experience of emotions,” Clinical psychology review, vol. 17, no. 7, pp. 791–821,1997.
3. B. Paul, S. Bera, T. Dey, and S. Phadikar, “Machine learning approach of speech emotions recognition using feature fusion technique,” Multimedia Tools and Applications, vol. 83, no. 3, pp. 8663–8688, 2024.
4. S. Sultana, M. Z. Iqbal, M. R. Selim, M. M. Rashid, and M. S. Rahman, “Bangla speech emotion recognition and cross-lingual study using deep cnn and blstm networks,” IEEE Access, vol. 10, pp. 564–578, 2021.
5. M. M. Billah, L. Sarker, M. Akhand, M. S. Kamal, et al., “Emotion recognition with intensity level from bangla speech using feature transformation and cascaded deep learning model,” International Journal of Advanced Computer Science & Applications, vol. 15, no. 4, 2024.


Thank You!
ICCIT
COX'S BAZAR
28th International
Conference on Computer
and Information Technology
2025